Fuzzy Structural Classification Methods

Mika Sato-Ilic[1] and Tomoyuki Kuwata[2]

[1] Faculty of Systems and Information Engineering, University of Tsukuba, Tsukuba, Ibaraki 305-8573, Japan
mika@risk.tsukuba.ac.jp

[2] Faculty of Systems and Information Engineering, University of Tsukuba, Tsukuba, Ibaraki 305-8573, Japan
kuwata10@sk.tsukuba.ac.jp

Abstract. This paper presents several fuzzy clustering methods based on self-organized similarity (or dissimilarity). Self-organized similarity (or dissimilarity) has been proposed in order to consider not only the similarity (or dissimilarity) between a pair of objects but also the similarity (or dissimilarity) between the classification structures of objects. Depending on how the similarity (or dissimilarity) of the classification structures cope with the fuzzy clustering methods, the results will be different from each other. This paper discusses this difference and shows several numerical examples.

1 Introduction

Recently, self-organized techniques were proposed and many algorithms have been developed and applied to many application areas. Clustering methods are no exception and much research based on this idea has been proposed [4], [8].

We also have proposed self-organized fuzzy clustering methods under an assumption that similar objects have similar classification structures [6], [7]. We have defined self-organized similarity and dissimilarity using weights which show the degree of similarity or dissimilarity between a pair of fuzzy classification structures. Given the self-organized similarity (or dissimilarity), it is known that the proposed methods tend to give clearer results. In this paper, we discuss several clustering methods based on self-organized similarity or dissimilarity and the features of these methods. These methods use the fuzzy clustering model [5] and the FANNY algorithm [3].

2 Additive Fuzzy Clustering Model

A fuzzy clustering model [5] has been proposed as follows:

$$s_{ij} = \sum_{k=1}^{K} \sum_{l=1}^{K} w_{kl} u_{ik} u_{jl} + \varepsilon_{ij}, \tag{1}$$

B. Gabrys, R.J. Howlett, and L.C. Jain (Eds.): KES 2006, Part II, LNAI 4252, pp. 102–109, 2006.

under the following conditions:

$$u_{ik} \in [0,1],\ \forall i,k;\ \sum_{k=1}^{K} u_{ik} = 1,\ \forall i. \tag{2}$$

In this model, u_{ik} shows degree of belongingness of an object i to a cluster k. The weight w_{kl} is considered to be a quantity which shows the asymmetric similarity between a pair of clusters. That is, we assume that the asymmetry of the similarity between the objects is caused by the asymmetry of the similarity between the clusters. We assume the following condition:

$$0 \leq w_{kl} \leq 1. \tag{3}$$

s_{ij} shows the similarity between objects i and j. ε_{ij} is an error. K shows the number of clusters and n is the number of objects. The purpose of model (1) is to find $U = (u_{ik})$ and $W = (w_{kl})$ which minimize the following sum of squares error η^2 under the conditions (2) and (3),

$$\eta^2 = \sum_{i \neq j=1}^{n} (s_{ij} - \sum_{k=1}^{K}\sum_{l=1}^{K} w_{kl} u_{ik} u_{jl})^2. \tag{4}$$

3 FANNY Algorithm

Fuzzy c-means (FCM) [1] is one of the methods of fuzzy clustering. FCM is a method which minimizes the weighted within-class sum of squares:

$$J(U, \boldsymbol{v}_1, \cdots, \boldsymbol{v}_K) = \sum_{i=1}^{n}\sum_{k=1}^{K} (u_{ik})^m\, d^2(\boldsymbol{x}_i, \boldsymbol{v}_k), \tag{5}$$

where $\boldsymbol{v}_k = (v_{ka}),\ k = 1, \cdots, K,\ a = 1, \cdots, p$ denotes the values of the centroid of a cluster k, $\boldsymbol{x}_i = (x_{ia}),\ i = 1, \cdots, n,\ a = 1, \cdots, p$ is i-th object with respect to p variables, and $d^2(\boldsymbol{x}_i, \boldsymbol{v}_k)$ is the square Euclidean distance between $\boldsymbol{x}_i$ and $\boldsymbol{v}_k$. The exponent m which determines the degree of fuzziness of the clustering is chosen from $(1, \infty)$ in advance. The purpose of this is to obtain the solutions U and $\boldsymbol{v}_1, \cdots, \boldsymbol{v}_K$ which minimize equation (5). The minimizer of equation (5) is shown as:

$$J(U) = \sum_{k=1}^{K} \left(\sum_{i=1}^{n}\sum_{j=1}^{n} ((u_{ik})^m (u_{jk})^m d_{ij}) / (2 \sum_{s=1}^{n} (u_{sk})^m) \right), \tag{6}$$

where $d_{ij} = d(\boldsymbol{x}_i, \boldsymbol{x}_j)$. The equation (6) is the objective function of the relational fuzzy c-means [2]. When $m = 2$, equation (6) is the objective function of the FANNY algorithm [3].

4 Self-organized Clustering Methods

4.1 Self-organized Clustering Methods Based on Additive Fuzzy Clustering Model

The method consists of the following three steps:

(Step 1) Apply the similarity data for model (1). If the similarity data is symmetric, we can obtain the symmetric matrix for $W = (w_{kl})$. Obtain the solutions $\hat{U} = (\hat{u}_{ik})$ and $\hat{W} = (\hat{w}_{kl})$.

(Step 2) Using the obtained $\hat{U}$, recalculate the following similarity:

$$\tilde{s}_{ij} = \frac{1}{\displaystyle\sum_{k=1}^{K}(\hat{u}_{ik} - \hat{u}_{jk})^2} s_{ij}, \quad i, j = 1, \cdots, n. \tag{7}$$

Using $\tilde{s}_{ij}$, go back to Step 1 and obtain a new result for $\tilde{U}$ and $\tilde{W}$.

(Step 3) Evaluate the fitness shown in equation (4) using $\tilde{U}$ and $\tilde{W}$ and compare with the fitness obtained by using $\hat{U}$ and $\hat{W}$. If the fitness using $\tilde{U}$ and $\tilde{W}$ is smaller than the fitness using $\hat{U}$ and $\hat{W}$, then replace $\hat{U}$ and $\hat{W}$ by $\tilde{U}$ and $\tilde{W}$, and repeat Steps 1 to 3. If the fitness using $\tilde{U}$ and $\tilde{W}$, and the difference between the fitness using $\hat{U}$ and $\hat{W}$ and the fitness using $\tilde{U}$ and $\tilde{W}$ are sufficiently small, then stop. Otherwise, repeat Steps 1 to 3, using new initial values for U and W.

Equation (7) shows that if $\hat{u}_{ik}$ and $\hat{u}_{jk}$ are similar to each other, then the similarity between objects i and j becomes larger. This similarity is self organizing according to the degree of belongingness for each of the clusters obtained in each iteration. In equation (7), we can rewrite $\tilde{s}_{ij}$ when we assume W is a unit matrix I in equation (1) as follows:

$$\tilde{s}_{ij} = \frac{\displaystyle\sum_{k=1}^{K}\hat{u}_{ik}\hat{u}_{jk}}{\displaystyle\sum_{k=1}^{K}(\hat{u}_{ik} - \hat{u}_{jk})^2}, \quad i, j = 1, \cdots, n. \tag{8}$$

If $W = I$ and $\sum_{k=1}^{K}(\hat{u}_{ik} - \hat{u}_{jk})^2$ is constant, then equation (8) is essentially the same as equation (1). $\sum_{k=1}^{K}(\hat{u}_{ik} - \hat{u}_{jk})^2$ has a bias because of condition (2). This method does not rectify the bias, it rather uses the bias to its advantage. Since this bias tends to make a cluster in which the objects do not have clear classification structures, we can obtain the defuzzified result, while ensuring the features of the fuzzy clustering result. Figure 1 shows the bias. We assume two degrees of belongingness, $\boldsymbol{u}_1 = (u_{11}, u_{12})$ and $\boldsymbol{u}_2 = (u_{21}, u_{22})$ corresponded to two objects 1 and 2. When we fix $\boldsymbol{u}_2 = (0.5, 0.5)$, the solid line shows the value of

$$\sum_{k=1}^{2} u_{1k}u_{2k}. \tag{9}$$

The abscissa shows values of u_{11}. Due to the condition (2), it is enough to only determine the value of u_{11}. The dotted line shows the square Euclidean distance between the degree of the belongingness for objects 1 and 2 as follows:

$$\sum_{k=1}^{2}(u_{1k} - u_{2k})^2. \tag{10}$$

From figure 1, we can see that a clearer result will make a large distance from the fixed point $(0.5, 0.5)$, even if the value of the inner product is the same. For example, if the values of u_{11} are 0.8 and 0.7, the values of equation (9) are the same for both 0.8 and 0.7. However, the values of equation (10) are different from each other. The classification structure is clearer, that is when $(u_{11}, u_{12}) = (0.8, 0.2)$ has a larger distance when compared with the case of the classification structure when $(u_{11}, u_{12}) = (0.7, 0.3)$. This is caused by a bias under the condition of $\sum_{k=1}^{K} u_{ik} = 1$.

Figure 2 shows the situation of the bias when the number of clusters is 3. Each axis shows each cluster. The solutions for objects 1, 2, 3, and 4 are on the triangle in figure 2. In this figure, the distance between $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ are larger than the distance between $\boldsymbol{u}_3$ and $\boldsymbol{u}_4$, even if the angle θ are almost the same.

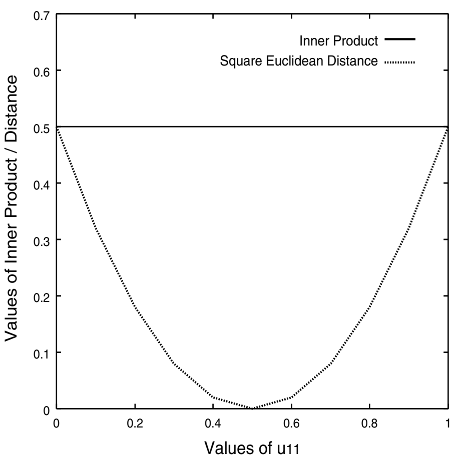

Fig. 1. Difference between Inner Product and Square Euclidean Distance

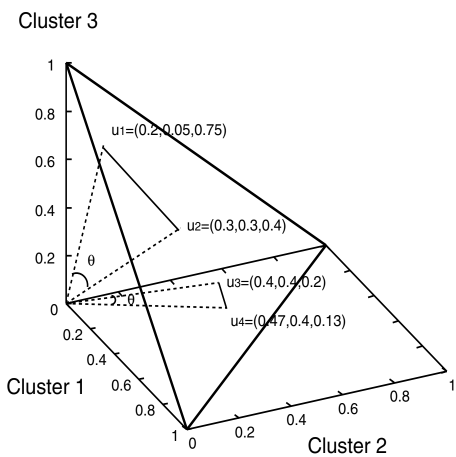

Fig. 2. Bias of the Distance

If we ignore the self-organization, then we can treat equation (8) as the following model.

$$s_{ij} = \frac{\sum_{k=1}^{K} u_{ik}u_{jk}}{\sum_{k=1}^{K}(u_{ik} - u_{jk})^2} + \varepsilon_{ij}, \quad i, j = 1, \cdots, n. \tag{11}$$

4.2 Self-organized Clustering Methods Based on the FANNY Algorithm

Next, we use the FANNY algorithm shown in equation (6). We define the self-organized dissimilarity as

$$\tilde{d}_{ij} = \sqrt{\sum_{k=1}^{K} (\hat{u}_{ik} - \hat{u}_{jk})^2 d_{ij}}, \quad i, j = 1, \cdots, n. \tag{12}$$

Here $\hat{u}_{ik}$ is degree of belongingness of an object i to a cluster k obtained which minimize equation (6) when $m = 2$. The algorithm is as follows:

(Step 1) Apply the dissimilarity data for the equation (6) when $m = 2$. Obtain the solution $\hat{U} = (\hat{u}_{ik})$.

(Step 2) Using the obtained $\hat{U}$, recalculate the following dissimilarity:

$$\tilde{d}_{ij} = \sqrt{\sum_{k=1}^{K} (\hat{u}_{ik} - \hat{u}_{jk})^2 d_{ij}}, \quad i, j = 1, \cdots, n.$$

Using $\tilde{d}_{ij}$, go back to Step 1 and obtain the new result for $\tilde{U}$.

(Step 3) Calculate the value of $\|\hat{U} - \tilde{U}\|$, where $\|\cdot\|$ shows the norm of matrix.

(Step 4) If $\|\hat{U} - \tilde{U}\| < \varepsilon$, then stop, or otherwise repeat Steps 1 to 4.

In equation (12), we assume $\hat{\boldsymbol{u}}_i = \hat{\boldsymbol{u}}_j, \quad (\hat{\boldsymbol{u}}_i = (\hat{u}_{i1}, \cdots, \hat{u}_{iK}), \ i = 1, \cdots, n)$. Then even if $d_{ij} \neq 0$, $\tilde{d}_{ij} = 0$. If $\hat{u}_{ik} \in \{0, 1\}$, then the possibility that the above situation occurs becomes larger. That is we stress the dissimilarity of the classification structures rather than the dissimilarity of the objects.

Considering the dissimilarity of degree of the belongingness for each object, we transform the data as follows:

$$\tilde{X} = (U|X).$$

Here $\tilde{X}$ is a $n \times (K + p)$ matrix. Using this matrix $\tilde{X}$, we apply the conventional FCM or FANNY algorithm. Since the distance between objects i and j is

$$\tilde{\tilde{d}}_{ij} = \sqrt{\sum_{k=1}^{K} (\hat{u}_{ik} - \hat{u}_{jk})^2 + \sum_{a=1}^{p} (x_{ia} - x_{ja})^2},$$

the clustering is considering not only the dissimilarity of objects but also the dissimilarity of degree of belongingness. In this case, even if $\boldsymbol{u}_i = \boldsymbol{u}_j$, then $\tilde{\tilde{d}}_{ij} \neq 0$, unless $\boldsymbol{x}_i = \boldsymbol{x}_j$. This method is not self-organized using the weights of dissimilarity of the classification structures.

5 Numerical Examples

We now show an example which uses real observations. Landsat data observed over the Kushiro marsh-land is used. The value of the data shows the amount of reflected light from the ground with respect to 6 kinds of light for 75 pixels. We get data from mountain area, river area, and city area. The 1st to the 25th pixels show mountain area, the 26th to the 50th are river area, and 51st to the 75th show the city area. The results are shown in figures 3 and 4.

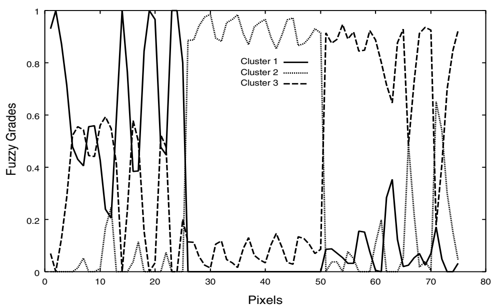

Fig. 3. Result of Landsat Data using the Fuzzy Clustering Model

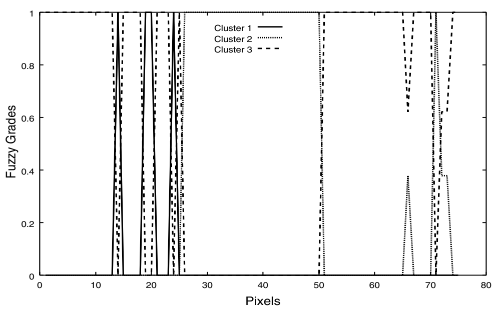

Fig. 4. Result of Landsat Data using the Self-Organized Fuzzy Clustering Method based on Equation (7)

Figure 3 shows the result using the conventional fuzzy clustering model shown in equation (1) and figure 4 is the result of the self-organized fuzzy clustering method using equation (7). In these figures, the abscissa shows each pixel and the ordinate shows the degree of belongingness for each cluster. The number of clusters is 3. From these results, we see that the result of the proposed method obtains a clearer result when compared with the result shown in figure 3.

Table 1. Comparison of the Fitness

	$\hat{\eta}^2$	$\tilde{\eta}^2$
Fitness	0.14	0.07

Table 1 shows the comparison of the fitness for both the methods. From this table, we see that a better solution is obtained by using the self-organized fuzzy clustering method using equation (7).
In table 1, $\hat{\eta}^2$ and $\tilde{\eta}^2$ are as follows:

$$\hat{\eta}^2 = \frac{\sum_{i \neq j=1}^{n} (s_{ij} - \sum_{k=1}^{K}\sum_{l=1}^{K} \hat{w}_{kl}\hat{u}_{ik}\hat{u}_{jl})^2}{\sum_{i \neq j=1}^{n} (s_{ij} - \bar{s}_{ij})^2}, \quad \bar{s}_{ij} = \frac{\sum_{i \neq j=1}^{n} s_{ij}}{n(n-1)}.$$

$$\tilde{\eta}^2 = \frac{\sum_{i \neq j=1}^{n} (\tilde{s}_{ij} - \sum_{k=1}^{K}\sum_{l=1}^{K} \tilde{w}_{kl}\tilde{u}_{ik}\tilde{u}_{jl})^2}{\sum_{i \neq j=1}^{n} (\tilde{s}_{ij} - \bar{\tilde{s}}_{ij})^2}, \quad \bar{\tilde{s}}_{ij} = \frac{\sum_{i \neq j=1}^{n} \tilde{s}_{ij}}{n(n-1)}.$$

We use the Kushiro marsh-land data for 4087 pixels over 6 kinds of light. We apply the self-organized fuzzy clustering discussed in section 4.2. The number of clusters is assumed to be 3. Figures 5 to 7 show the results of degree of belongingness for the FANNY method, our proposed method when one iteration is used, and our proposed method until it is convergent, respectively. These figures show the results of cluster 1. In these figures each color shows the range of degree of belongingness to each cluster. Red shows $0.8 < u_{ik} \leq 1.0$, yellow shows $0.6 < u_{ik} \leq 0.8$, light blue shows $0.4 < u_{ik} \leq 0.6$, blue is $0.2 < u_{ik} \leq 0.4$, and white is $0 < u_{ik} \leq 0.2$. From these figures, we can see that the degree of belongingness for the clusters is going to create a crisper situation.

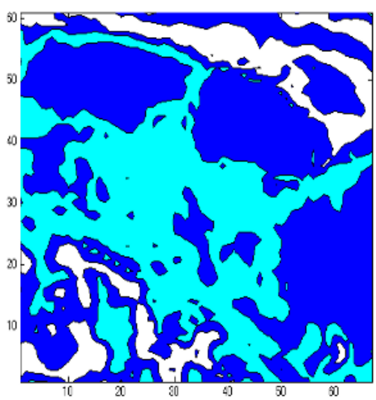

Fig. 5. Result of FANNY for Cluster 1

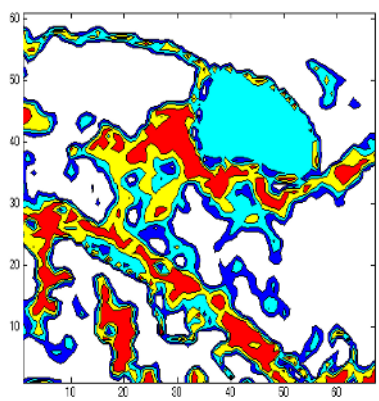

Fig. 6. Result of Self-Organized Method for Cluster 1 (after one iteration)

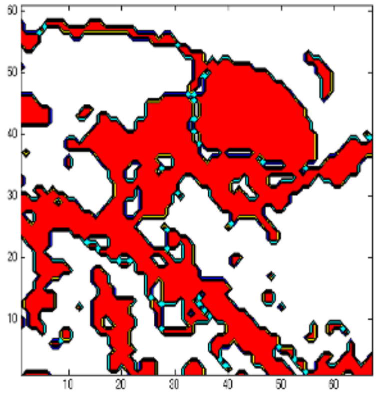

Fig. 7. Result of Self-Organized Method for Cluster 1 (after the convergence)

6 Conclusion

We discussed several methods using self-organized similarity (or dissimilarity) for fuzzy clustering methods. The concept of "self-organization" refers to the situation where the similarity of objects is affected by the similarity of classification structures of objects in each iteration. We consider two spaces. One is the observation space of objects and the other is a space of degree of belongingness of objects to clusters which show the classification structures corresponding to each object. As a result, we could obtain a clearer result and it ensures applicability as a defuzzification method.

References

1. Bezdek, J. C.: Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum Press. (1987)
2. Hathaway, R. J., Davenport, J. W., and Bezdek, J. C.: Relational Duals of the C-Means Clustering Algorithms. Pattern Recognition. **22** (1989) 205–212
3. Kaufman L., Rousseeuw, P. J.: Finding Groups in Data. John Wiley & Sons. (1990)
4. Kohonen, T.: Self-Organized Formation of Topologically Correct Feature Maps. Biol. Cybern. **43** (1982) 59–69
5. Sato, M., Sato, Y., and Jain, L. C.: Fuzzy Clustering Models and Applications. Springer. (1997)
6. Sato-Ilic, M.: Self Organized Fuzzy Clustering. Intelligent Engineering Systems through Artificial Neural Networks. **14** (2004) 579–584
7. Sato-Ilic, M., Kuwata, T.: On Fuzzy Clustering based Self-Organized Methods. FUZZ-IEEE 2005. (2005) 973–978
8. Van Hulle, M. M.: Faithful Representations and Topographic Maps -From Distortion to Information based Self-Organization. John Wiley & Sons. (2000)